基于距离聚类与K-means动态聚类的棉田土壤养分评价研究

范向龙[1,2]， 吕 新[1,2]， 张 泽[1,2]， 高 攀[2,3]，
张 强[1,2]， 印彩霞[1,2]， 易 翔[1,2]

（1. 石河子大学农学院，新疆生产建设兵团绿洲生态农业重点实验室，新疆 石河子 832003；
2. 新疆兵团农业大数据国家地方联合工程研究中心，新疆 石河子 832003；
3. 石河子大学信息科学与技术学院，新疆 石河子 832003）

摘 要：为了解棉田土壤养分状况及确定养分丰缺程度，研究采用主成分结合距离聚类、K-means动态聚类以及土壤养分综合评价方法对新疆棉田土壤进行分析评价。结果表明：(1) 在主成分分析中，有效铜、有效锰和碱解氮起主要作用，有效铜含量为1.82 mg·kg^{-1}，丰缺评价状况属于高等水平，有效锰含量为11.36 mg·kg^{-1}，属于较低水平，碱解氮含量为122.07 mg·kg^{-1}，属于高等水平，土壤速效磷、速效钾、有效锌和有效铁含量较低，土壤养分含量分布不均匀。在距离聚类和K-means动态聚类中，有机质、碱解氮、有效锰含量较低，其余养分含量较高。在距离聚类中，土壤各类养分可表示为：第Ⅰ类>第Ⅴ类>第Ⅳ类>第Ⅱ类>第Ⅲ类，而在K-means动态聚类中可以表示为：第Ⅲ类>第Ⅰ类>第Ⅴ类>第Ⅱ类>第Ⅳ类。(2) 在土壤综合肥力指数评价值中（IFI），1连和16连的等级高；2连、3连、4连、6连、15连、19连和二监区的等级较高；8连、9连、10连、11连、12连、17连、18连和20连在中等水平。5连、7连、一监区和农市站的等级较低。K-means动态聚类比距离聚类分类效果好，可以更加科学合理、准确有效地对土壤养分进行综合评价。

关键词：土壤养分；距离聚类；K-means动态聚类；综合评价值

土壤对作物生长起着决定性的作用，土壤肥力是土壤物理结构及化学性质的综合表现[1]。土壤养分是评价土壤肥力的主要指标[2]，决定着农作物产量的高低[3]。人们经常用土壤养分的均值来代表整个田块的土壤养分状况，按单个养分含量进行施肥，此种施肥模式往往达不到按需施肥的目的，反而降低了肥料的利用效率。因此，综合土壤养分多个指标对其进行分析判别是非常有必要的，既可以客观评价土壤养分肥力状况，又可提高肥料的利用效率，增加经济效益。

戴余波等[4]运用主成分分析法，将耕地土壤养分分为4个等级，对于肥力较差的土壤制定相应的施肥方案；刘少春等[5]采用主成分分析对蔗田土壤进行评价，分析发现有机质不仅与土壤的氮、磷有关，还与微量元素铜、锰、锌有着密切的关系，并得出有机质可以作为反映土壤肥力的一个重要综合指标；李世瑶等[6]采用主成分分析将各评价指标的加权和作为灌水质量的综合主成分指标，发现综合主成分能更好的反应灌水质量变异信息；安云娜等[7]运用模糊数学和偏相关分析法得出土壤养分的综合指标值，并提出了不同施肥意见及土地改良措施；唐佐芯等[8]通过主成分与聚类方法分析发现地形、海拔和pH是影响植烟土壤分类的主要环境因素；郑琦等[9]采用模糊综合指数法（FCE）、土壤综合质量指数法（SQI）及内梅罗综合污染指数法（N-PI）分别对棉田土壤进行分析评价，得出SQI评价法

收稿日期：2020-08-28； **修订日期：**2020-10-02
基金项目：兵团重点领域创新团队项目（2018CB004）；兵团国际合作计划项目（2018BC009）；国家博士后面上项目（2017M623282）；石河子大学创新发展专项（CXFZ201903）；新疆兵团棉花生产大数据关键技术及农业大数据平台研发应用（2018AA004）
作者简介：范向龙（1992-），男，博士研究生，研究方向为农业信息技术. E-mail: 1574468658@qq.com
通讯作者：吕新. E-mail: lxshz@126.com

比FCE和N-PI评价法效果好。

综上所述，已有研究主要通过主成分或聚类等单一方法对土壤养分进行分析，而运用主成分与聚类方法相结合在土壤养分的研究中较少，主成分分析的特征是对信息贡献影响力综合评价，结合聚类分析法，可以更精准、更加全面真实的反应土壤养分综合状况。因此，本研究对新疆生产建设兵团农二师33团的土壤进行相关分析，通过主成分结合距离聚类以及K-means动态聚类的方法对多个土壤养分指标进行分级判别与评价，建立其更加合理有效的土壤养分评价方法，以期为科学施肥提供指导。

1　研究区概况与方法

1.1　研究区概况

新疆生产建设兵团农二师33团(87°03′30″E，40°40′00″~40°53′52″N)位于塔里木河下游，地处塔克拉玛干大沙漠腹地、东北边缘的孔雀河和西南方向的塔里木河之间[10]。研究区日照时间比较长，光热资源丰富，无霜期平均210 d，年平均降水量58.6 mm，年最大蒸发量为2788.2 mm，降水少而蒸发大，冬寒夏热，昼夜温差悬殊，属典型的暖温带大陆性干旱气候，对农业的发展有重要的影响，通过多年的发展，已成为南疆棉花生产的重要组成部分[11]。

1.2　数据来源及指标测定

采用2013—2015年的土壤养分数据从新疆生产建设兵团农二师33团获取，每个土壤样品选择五点取样法进行样品采集，试验区土地均匀平整，土壤取样深度为0~20 cm，每个采样点采样完成后用GPS进行定位(图1)。采样指标包括有机质(OM)、碱解氮(AN)、速效磷(AP)、速效钾(AK)、有效铜(ACu)、有效锌(AZn)、有效锰(AMn)、有效铁(AFe)，其中有机质采用重铬酸钾容量法测定(稀释热法)，碱解氮采用碱解扩散法测定，速效磷采用碳酸氢钠提取—钼锑抗比色法测定；速效钾采用乙酸铵浸提—火焰光度计法测定；有效铁、锰、铜、锌均采用DTPA浸提—电感耦合等离子体质谱法测定[12]。每个土样测定重复3次，结果取其平均值，最后对其进行等级划分[13-14]。

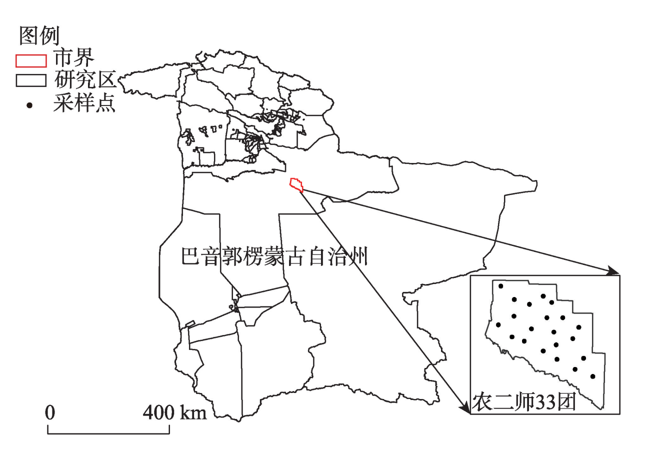


图1　农二师33团位置及采样点分布

Fig. 1　Distribution map of position and sampling points in 33 Mission of Agriculture Division 2

1.3　研究方法

1.3.1　主成分分析　由于土壤原始数据样本比较大，所得到的统计数据信息在一定程度上会出现重叠现象，变量过多会出现数据的复杂性[15]。主成分分析可有效解决以上缺陷，其主要作用是用少量指标来反映大量原始变量的信息[16]。

1.3.2　距离聚类　距离聚类在数据挖掘中属于无监督学习[17]，先将数据集划分为很多子集，让同一集合内的数据相似性增高，而不同集合间的数据差异变得更明显[18]。本文采用欧式距离对原始数据进行处理[19]，其优点是新增的对象不会涉及两个对象间的任何距离，计算公式：

$$d(x, y)=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2} \tag{1}$$

式中：x、y为数据对象；n为数据对象的维度；x_i、y_i为x和y的第i个属性值。

1.3.3　K-means聚类　K-means聚类属于无监督学习的一种聚类方法[20]。主要针对观测到的数据，根据给定的准则发现具有的共同点，在数据集中寻找“群”。K-means聚类以距离作为数据对象间相似性度量的标准[21-23]，其原理是先从所有数据集中随机选取初始聚类中心，计算剩余数据与聚类中心间的距离，找出离目标数据最近的聚类中心，并将数据对象分配到聚类中心所对应的簇中。然后将每个簇中的数据均值作为新的中心，进行下一次迭代，直到聚类中心不再变化为止[24]，计算公式如下：

$$d(X, C_i)=\sqrt{\sum_{j=1}^{m}(X_j-C_{ij})^2} \tag{2}$$

式中：X为数据对象；C_i为第i个聚类中心；m为数据对象的维度；X_j、C_{ij}为X和C_i的第j个属性值。

1.3.4　土壤肥力指数　土壤肥力指数是评价土壤肥力常用的方法[25]。土壤肥力指数的计算方法有：直

接叠加法、权重加权求和法、综合评价模型等。本文主要采用综合肥力指数评价模型(IFI)[26]。土壤综合肥力指标值(IFI)划分为5个等级,即:低(IFI＜0.2)、较低0.2≤IFI＜0.4)、中(0.4≤IFI＜0.6)、较高(0.6≤IFI＜0.8)和高(IFI≥0.8)。IFI能够全面反映土壤养分状况,取值范围在0~1之间,越接近1,说明土壤肥力越好[27]。计算公式如下:

$$IFI=\sum_{i=1}^{n}Q_iY_i \quad (3)$$

式中:n为指标个数;Q_i和Y_i分别为第i种肥力指标的隶属度和相应的权重。

通过计算8个土壤养分指标的公共因子方差,将公共因子方差进行归一化,结果作为各项肥力指标的权重。通过模糊数学中的隶属度函数法对数据进行了标准化处理,以消除量纲的影响[28]。由于土壤养分因子变化具有连续性,所以在评价指标时可采用连续性的升型隶属度函数[29]。

2 结果与分析

2.1 土壤养分主要特性

通常用变异系数(CV)来反映土壤养分的空间分布状况,当$CV<0.1$时为弱变异,CV在0.1~1.0时为中等变异,$CV>1.0$时为强变异[30-31],土壤养分特性由表1可知,研究区中OM、AN变异系数最低,为0.16,表明AN在土壤中分配较集中,而AZn变异系数最高,为0.61,AZn在土壤中分配比较离散,剩余各养分指标均属于中等变异。

2.2 土壤养分特征元素的提取

本文主要以8个土壤养分指标为基础数据,通过主成分分析做相关矩阵及贡献率,找出土壤中少量所具备代表性的综合指标,以较少的土壤信息变量为基准,来反映出土壤原来的总体变量信息(表2)。由表2可知,前3个主成分特征值均大于1,累计贡献率为88.69%,即涵盖了最初8个指标的原始数据信息,表明前3项主成分能够代表最初的8个指标原始数据信息。

由土壤养分主成分旋转后的载荷系数可知(表3),第一主成分与ACu高度正相关,主成分载荷为0.966,说明第一主成分是ACu的综合反映;第二主成分中,AMn的主成分载荷最高为0.812,说明第二主成分主要反映的是AMn对土壤养分的供给情况;第三主成分中,AN的主成分载荷最高为0.964,表明第三主成分是对AN含量的描述。因为总方差近88.69%的贡献率来自于这3个因素,由此认为ACu、AMn、AN是农二师33团土壤的特征元素。

2.3 基于距离聚类的土壤养分分类评价

距离聚类分类结果见图2,根据土壤养分含量利用距离聚类将农二师33团划分为5种类型,第Ⅰ类有16连,与相近的地区土壤养分差异较大;第Ⅱ类有二监区;第Ⅲ类包括5连、12连、农市站和一监区;第Ⅳ类包括3连和4连;第Ⅴ类包括1连、2连、6连、7连、8连、9连、10连、11连、15连、17连、18连、19连、20连。

由土壤养分各元素含量比较可知(表4),AN、AP、AK、AFe含量在各类别间差异比较大,其余各土壤养分含量变化差异不大。OM含量在第Ⅰ类、第Ⅳ类和第Ⅴ类间及第Ⅱ类和第Ⅲ类间差异不显著;AN含量在第Ⅱ类与第Ⅴ类及第Ⅲ类与第Ⅳ类间差异不显著;AP含量在第Ⅰ类和第Ⅱ类及第Ⅳ类和第

表1 土壤养分各指标描述统计分析

Tab. 1 Statistical analysis of soil nutrient indicators

指标/($mg\cdot kg^{-1}$)	最大值	最小值	均值	极差	标准差	变异系数
OM	11220	6310	9370	4910	1510	0.16
AN	158.32	88.63	122.07	69.69	19.77	0.16
AP	26.39	10	20.36	16.39	4.44	0.21
AK	192.13	143.6	149.77	48.53	29.97	0.20
AFe	15.63	10.79	9.25	4.84	3.29	0.35
AMn	15.18	3.65	11.36	11.53	3.81	0.33
ACu	2.41	1.55	1.82	0.86	0.61	0.33
AZn	3.45	1.97	1.07	1.48	0.651	0.61

注:OM表示有机质;AN表示碱解氮;AP表示速效磷;AK表示速效钾;AFe表示有效铁;AMn表示有效锰;ACu表示有效铜;AZn表示有效锌。下同。

表2　研究区土壤养分PCA解释主变量

Tab. 2　PCA interpretation of soil nutrients in study area

主成分数	特征值	累计方差贡献率/%
1	4.260	53.252
2	1.757	75.212
3	1.078	88.690
4	0.578	95.914
5	0.174	98.084
6	0.117	99.553
7	0.032	99.956
8	0.003	100.000

表3　研究区土壤养分主成分旋转后的载荷矩阵

Tab. 3　Component load matrix of soil main nutrient after rotation in the study area

指标	PC1	PC2	PC3
OM	0.693	0.670	0.147
AN	−0.157	0.038	0.964
AP	0.319	0.809	0.261
AK	0.517	0.719	0.377
AFe	0.593	0.164	0.747
AMn	−0.220	0.812	−0.268
ACu	0.966	0.084	0.028
AZn	0.538	0.025	0.700

Ⅴ类之间差异不显著，第Ⅲ类与其余各类别间AP含量差异显著；AK含量除了第Ⅲ类与其余各类间差异显著，剩下的各类别间AK含量差异不显著；AFe含量除了第Ⅰ类、第Ⅱ类和第Ⅳ类间差异不显著外，其余各类间AFe含量差异显著；AMn含量在第Ⅱ类和第Ⅳ类及第Ⅲ类和第Ⅴ类间差异不显著；ACu含量除了第Ⅰ类与其余各类间差异显著外，剩余各类别间ACu含量差异不显著；AZn含量在第Ⅰ类和第Ⅳ类及第Ⅱ类、第Ⅲ类和第Ⅴ类间差异不显著，其余各类别间差异显著；土壤各类别养分总体状况可以表示为：第Ⅰ类>第Ⅴ类>第Ⅳ类>第Ⅱ类>第Ⅲ类。

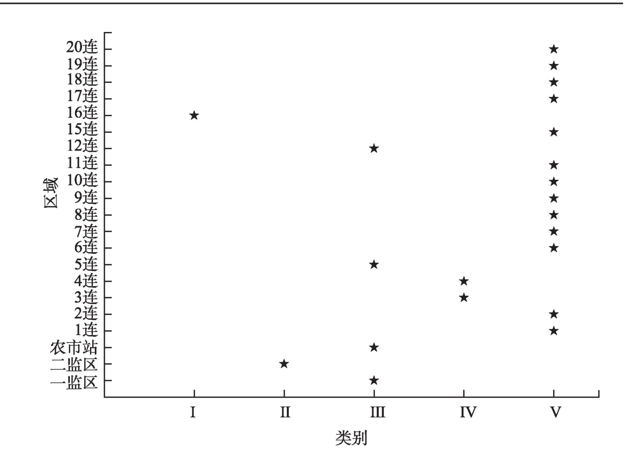


图2　土壤养分距离分类

Fig. 2　Soil nutrient distance clustering

2.4　基于K-means动态聚类的土壤养分分类评价

通过K-means动态聚类方法将农二师33团土壤养分分为5种类型(图3)，第Ⅰ类包括1连、3连、4连、15连；第Ⅱ类包括7连、8连、9连、10连、11连、17连、18连、19连、20连；第Ⅲ类包括2连、6连、16连；第Ⅳ类包括5连、12连、一监区和农市站；第Ⅴ类只有二监区。

从各类别土壤养分含量可知(表5)，土壤养分含量变化差异比较大的是AP、AK、AFe。OM含量中第Ⅳ类和第Ⅴ类差异不显著，第Ⅱ类和第Ⅲ类OM含量差异显著；AN含量各类别间差异不显著；AP含量在第Ⅰ类和第Ⅴ类及第Ⅱ类和第Ⅲ类间差异不显著，第Ⅲ类与其余各类别间AP含量差异显著；AK含量在第Ⅰ类、第Ⅱ类和第Ⅴ类间含量差异不显著，第Ⅲ类和第Ⅳ类与其余各类别间AK含量差异显著；AFe含量在第Ⅰ类、第Ⅲ类和第Ⅴ类间差异不显著；AMn含量在第Ⅰ类和第Ⅴ类及第Ⅱ类、第Ⅲ类和第Ⅳ类间差异不显著；ACu含量除了第Ⅲ类与其余各类别间差异显著外，其余各类别间ACu含量

表4　各类别土壤养分元素含量比较

Tab. 4　Comparison of soil nutrient elements in each category

类别	OM	AN	AP	AK	AFe	AMn	ACu	AZn
Ⅰ	1.03a	86.37a	20.08b	192.31a	15.72a	3.45ab	2.42a	3.21a
Ⅱ	0.67b	73.77b	17.87b	170.96a	13.96a	2.43b	1.40b	1.36b
Ⅲ	0.76b	61.38c	11.18c	111.26b	5.91c	4.57a	1.79b	1.66b
Ⅳ	0.98a	57.85c	28.68a	176.75a	14.5a	2.88b	1.86b	2.68a
Ⅴ	1.02a	70.18b	20.89a	192.43a	8.88b	4.44a	1.60b	1.77b

注：同列数据不同小写字母表示不同土壤类别间差异显著水平($P<0.05$)。

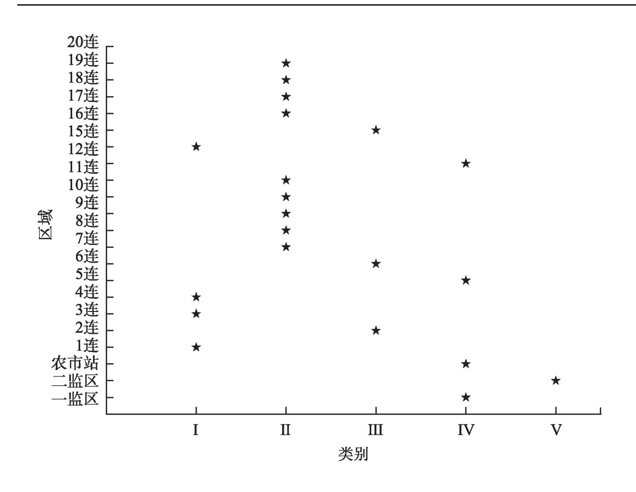


图3 K-means动态聚类

Fig. 3 K-means dynamic clustering

差异不显著；AZn含量在第Ⅱ类和第Ⅴ类间差异不显著。土壤各类别养分总体状况可以表示为：第Ⅲ类>第Ⅰ类>第Ⅴ类>第Ⅱ类>第Ⅳ类。

2.5 土壤肥力权重及隶属度值的确定

通过对8个土壤养分指标进行主成分分析(表2)，经计算得出土壤养分各指标的公因子方差和权重(表6)，AFe、AMn和AN的权重高，分别为0.5327、0.2196、0.1348，AK、OM和AZn的权重较低，分别为0.0722、0.0217、0.0146，AP和ACu的权重最低，为0.0003和0.0039。

由土壤各指标隶属度统计可知(表7)，OM、AN、AP、AFe和AMn的变异系数在0.409~0.491；AK和AZn的变异系数0.305和0.371；ACu的变异系数在0.511。各指标的隶属度值总体变异系数较小，因此，研究区8个土壤养分指标可全部作为土壤肥力质量敏感的参评指标。

2.6 土壤养分状况评价

从土壤的综合肥力指数可以看出(表8)，依据土壤综合肥力指标等级划分，2连、3连、4连、6连、15连、19连和二监区的IFI值在较高水平；8连、9连、10连、11连、12连、17连、18连和20连的IFI值在中等水平；5连、7连、一监区和农市站的IFI值在较低水平。

表5 各类别土壤养分含量

Tab. 5 Soil nutrient content of each category

类别	OM	AN	AP	AK	AFe	AMn	ACu	AZn
Ⅰ	1.03ab	62.18a	27.64a	166.49b	12.32a	2.96b	1.85b	2.02b
Ⅱ	1.00b	61.75a	17.96b	160.62b	7.67b	4.54a	1.63b	1.55c
Ⅲ	1.11a	71.87a	19.35b	192.69a	12.43a	4.05a	2.92a	3.01a
Ⅳ	0.70c	63.48a	10.46c	110.42c	6.76b	4.23a	1.73b	1.71ab
Ⅴ	0.65c	74.43a	18.36b	166.33b	13.5a	2.58b	1.34b	1.37c

表6 土壤各指标公因子方差及权重

Tab. 6 Variance and weight of common factors of soil indexes

	OM	AN	AP	AK	AFe	AMn	ACu	AZn
公因子方差	0.8943	0.9383	0.7335	0.9184	0.9971	0.9565	0.7483	0.8275
权重	0.0217	0.1348	0.0003	0.0722	0.5327	0.2196	0.0039	0.0146

表7 土壤各指标隶属度统计

Tab. 7 Statistics of membership degree of soil indexes

指标/($g\cdot kg^{-1}$)	最大值	最小值	均值	标准差	变异系数
OM	0.991	0.100	0.648	0.270	0.417
AN	0.992	0.098	0.549	0.243	0.442
AP	0.994	0.148	0.583	0.238	0.409
AK	0.921	0.098	0.619	0.189	0.305
AFe	0.994	0.097	0.472	0.232	0.491
AMn	0.974	0.102	0.549	0.257	0.468
ACu	0.836	0.065	0.328	0.167	0.511
AZn	0.921	0.100	0.366	0.135	0.371

表8 研究区土壤的综合肥力指数

Tab. 8 Comprehensive soil fertility index of the study area

区域	IFI	等级	区域	IFI	等级	区域	IFI	等级
1连	1.596	高	8连	0.473	中等	17连	0.533	中等
2连	0.628	较高	9连	0.483	中等	18连	0.591	中等
3连	0.733	较高	10连	0.513	中等	19连	0.618	较高
4连	0.704	较高	11连	0.513	中等	20连	0.552	中等
5连	0.225	较低	12连	0.451	中等	一监区	0.383	较低
6连	0.634	较高	15连	0.705	较高	二监区	0.765	较高
7连	0.356	较低	16连	0.922	高	农市站	0.292	较低

通过距离聚类及K-means动态聚类对新疆生产建设兵团农二师33团棉田土壤进行分类评价，不同的分类方法根据其距离中心及分类原理将土壤养分为5类，距离聚类和K-means动态聚类中第Ⅰ类与第Ⅲ类土壤养分含量整体高于其他几类，但是距离聚类中的第Ⅲ类与K-means动态聚类中的第Ⅳ类土壤养分含量相近，这是由于不同分类方法中各土壤养分之间以及养分离聚类中心点是不同的，导致其分类结果出现偏差。通过土壤肥力权重及隶属度值的计算，发现农二师33团土壤养分差异比较大，一方面可能因为新疆棉区主要分布在盆地周围，降水较少，并且土质疏松，土壤理化性质较差，导致土壤养分容易流失，另一方面可能与当地有机肥的施入有关。通过土壤肥力综合指数(IFI)进行等级划分，分析得出农二师33团各地区处于中等和高等水平的所占比例最大，处于低等水平所占比例少，说明土壤肥力高，适合农作物生长。将距离聚类和K-means动态聚类的分类结果与土壤肥力综合指数等级划分结果相比，分析发现K-means动态聚类的分类结果更加接近土壤肥力综合指数的分类结果，因此，K-means动态聚类在土壤养分的分类方法上其效果更好，后期可直接用主成分结合K-means动态聚类对土壤养分进行分析评价。

3 讨论

3.1 土壤养分机理分析

土壤养分是反映土壤肥力高低的关键指标之一，直接影响着作物的生长发育及产量的高低[32-34]，土壤养分的评价可以为土地合理使用、农作物种植、农田施肥、土壤污染、科学治理等提供建议。土壤养分是一个动态变化的过程。本文通过主成分分析得出农二师33团土壤养分主要以AN、ACu和AMn为主，其他元素含量较低，这与田立文等[35]的研究不一致，主要由于农二师33团农民种植过程中不施用微量元素肥料，并且长期用苦咸水来灌溉棉田，苦咸水中含有少量微量元素，以磷肥最为突出，但这与土壤中微量元素含量变化是否有关，有待进一步研究。新疆种植区域以盐碱地为主，微量元素随土壤pH的升高而减小，有效性也随之降低，土壤中微量元素的有效性大部分存在于矿物晶格中，不能被植物吸收利用。土壤养分全量仅能看储量指标，不能反映其有效性，微量元素的供给不足除了有可能是因微量元素的全量低之外，还有一种可能是土壤中微量元素有效性低。因此，合理调控当地土壤中微量元素的植物有效性是实现有效管理的关键，并且微量元素在土壤中的有效性与土壤-植物的相互作用紧密相关。

农二师33团氮肥含量多，磷钾肥含量少，由于当地灌溉主要选用滴灌技术，不同滴灌时段下的土壤养分变异不同，差异比较大，这与袁宇尧等[36]研究结果基本一致，当地施氮肥以提高棉花产量，氮肥过量棉花无法全部吸收，常年在土壤中积累，与张丹等[37]的研究结果不同，滴灌所带入土壤中的氮以硝态氮为主，滴灌棉田属于落干晒田，利于土壤通气性，促进作物生长。土壤氮一部分被作物所吸收，另一部分被微生物所吸收同化。碱性土壤能增加含磷化合物溶解度，导致土壤磷的有效性降低，频繁的干湿交替以及高温，会导致钾的释放。有机质与土壤微生物密切相关，含水量或温度过高会限制微生物的活动，影响有机质的分解和转化，有机质在土壤中的分解与转化比较复杂，其来源由动植物残体的数量和类型决定。当地灌水比较多，但每年将秸秆归还大田中，土壤有机碳与活性有机碳含

量会明显增高，所以有机质含量总体不是很低。土壤养分丰缺与长期连作、合理灌溉施肥、膜下滴灌技术等有着密切的关系。所以在今后土壤养分供给中，应因地制宜合理施肥，保证土壤养分的协调供应。

3.2 距离聚类与K-means动态聚类可行性分析比较

距离聚类与K-means动态聚类将土壤养分各划分为5类，在距离聚类中，第Ⅰ类土壤养分含量整体高于其他几类，而在K-means动态聚类中，第Ⅲ类土壤养分含量整体要高于其他类，各土壤养分之间以及养分离聚类中心点是不同的，导致其分类结果不同。通过土壤肥力权重及隶属度值的确定，发现农二师33团土壤养分差异比较明显，但土壤养分各元素分布比较集中，这与陈彦[38]的研究结果相同，新疆地域辽阔，土壤本身属性差异与不同棉区的成土母质有关，产棉区由于不同区域、气候、温度、地形等环境差异，在空间位置上比较分散，而且新疆以滴灌为主，所以使得土壤养分变异更加分散，不管是在水平方向还是垂直方向都会出现。

土壤肥力综合指数(IFI)能更加全面反映土壤养分状况，在其等级划分中，中等和高等水平所占比例最大，低等水平所占比例少，则表明研究区土壤肥力较好。距离聚类的分类结果与土壤养分综合评价值的分类结果差异比较大，而K-means动态聚类与土壤养分综合评价值的分类结果比较接近，因此，K-means动态聚类对土壤分类效果更好，这与孙亚洲[39]的研究结果一致，距离聚类是对原始数据进行优化分析，是根据各个养分指标间的空间距离而产生的聚类，其对数据挖掘比较浅，而K-means动态聚类则是对预处理后的数据进一步挖掘分析，是土壤养分各指标对其聚类中心点之间的相似度所产生的聚类，其结果更能反应土壤的真实情况。

4 结 论

通过对新疆生产建设兵团第二师33团棉田土壤养分进行主成分分析研究，筛选出具有代表性的土壤养分综合指标，通过不同聚类算法以及土壤养分综合评价值对该区域棉田土壤养分进行分析评价，主要得出以下结论：

(1) 通过主成分分析得出，基于主成分分析方法，研究区棉田土壤有效铜、有效锰和碱解氮起主要作用，土壤速效磷、速效钾、有效锌和有效铁含量较低，土壤养分含量分布整体表现不均匀。

(2) 采用距离聚类与K-means动态聚类方法将研究区棉田土壤养分分为5种类型，不同分类方法对土壤养分等级划分结果也不同。在距离聚类中各类别土壤养分类型划分情况为：第Ⅰ类>第Ⅴ类>第Ⅳ类>第Ⅱ类>第Ⅲ类，而在K-means动态聚类中土壤养分等级表现为：第Ⅲ类>第Ⅰ类>第Ⅴ类>第Ⅱ类>第Ⅳ类。

(3) 通过土壤养分综合评价值IFI划分肥力等级的方法可知，该棉区土壤综合肥力指数等级处于中等及较高水平，土壤肥力状况良好。土壤养分综合评价值的分类结果与距离聚类的分类结果差异比较大，而与K-means动态聚类分类结果相近，因此，K-means动态聚类比距离聚类分类效果好。

参考文献(References)：

[1] 陈玉芹, 胡永亮, 张丽萍, 等. 基于主成分和聚类分析的德宏橡胶林土壤肥力评价[J]. 热带作物学报, 2019, 40(8): 1461-1467. [Chen Yuqin, Hu Yongliang, Zhang Liping, et al. Evaluation of soil fertility of rubber plantation in Dehong based on principal component and cluster analysis[J]. Chinese Journal of Tropical Crops, 2019, 40(8): 1461-1467.]

[2] 任艳芳, 何俊瑜, 张艳超, 等. 贵州省开阳茶园土壤养分状况与肥力质量评价[J]. 土壤, 2016, 48(4): 668-674. [Ren Yanfang, He Junyu, Zhang Yanchao, et al. Soil nutrient status and comprehensive evaluation of quality of soil fertility of tea garden in Kaiyang of Guizhou Province[J]. Soils, 2016, 48(4): 668-674.]

[3] 赵月玲, 林玉玲, 曹丽英, 等. 基于主成分分析和聚类分析的土壤养分特性研究[J]. 华南农业大学学报, 2013, 34(4): 484-488. [Zhao Yueling, Lin Yuling, Cao Liying, et al. A study of soil nutrients characteristics based on principal component and cluster analysis[J]. Journal of South China Agricultural University, 2013, 34(4): 484-488.]

[4] 戴余波, 张丽萍, 李国明, 等. 热带作物耕地土壤养分分析及肥力评价[J]. 现代农业科技, 2017, 52(18): 155-157. [Dai Yubo, Zhang Liping, Li Guoming, et al. Soil nutrient analysis and soil fertility evaluation of tropical crop land[J]. Modern Agricultural Science and Technology, 2017, 52(18): 155-157.]

[5] 刘少春, 张跃彬, 郭家文, 等. 基于养分丰缺分级的蔗田土壤肥力主成分综合分析[J]. 西南农业学报, 2016, 29(3): 611-617. [Liu Shaochun, Zhang Yuebin, Guo Jiawen, et al. Comprehensive analysis of main components of soil fertility of sugarcane field based on soil nutrient grades[J]. Southwest China Journal of Agricultural Sciences, 2016, 29(3): 611-617.]

[6] 李世瑶, 蔡焕杰, 陈新明. 基于主成分分析的畦灌质量评价[J]. 农业工程学报, 2013, 29(24): 86-93. [Li Shiyao, Cai Huanjie, Chen Xinming. Evaluation of border irrigation performance based on principal component analyses[J]. Transactions of the Chinese Society of Agricultural Engineering, 2013, 29(24): 86-93.]

[7] 安云娜, 黄义雄, 官紫玲. 福建东山岛土壤养分综合评价[J]. 安徽农业科学, 2007, 35(13): 3926-3927. [An Yunna, Huang Yixiong, Guan Ziling. Comprehensive valuation on soil nutrient in Dongshan island of Fujian[J]. Journal of Anhui Agricultural Sciences, 2007, 35(13): 3926-3927.]

[8] 唐佐芯, 李天璧, 陈泽斌, 等. 基于主成分分析和聚类分析的红河州植烟土壤特性研究[J]. 西南农业学报, 2019, 32(7): 1607-1613. [Tang Zuoxin, Li Tianbi, Chen Zebin, et al. Study of Honghe tobacco-planting soil nutrients characteristics based on principal component and cluster analysis[J]. Southwest China Journal of Agricultural Sciences, 2019, 32(7): 1607-1613.]

[9] 郑琦, 王海江, 董天宇, 等. 基于不同评价方法的绿洲棉田土壤质量综合评价[J]. 灌溉排水学报, 2019, 38(3): 90-98. [Zheng Qi, Wang Haijiang, Dong Tianyu, et al. Evaluating soil quality of the cotton fields in Oasis of Xinjiang using different methods[J]. Journal of Irrigation and Drainage, 2019, 38(3): 90-98.]

[10] 胡俊建. 守护沙漠绿洲——新疆兵团第二师三十三团生态文明建设纪实[J]. 中国农垦, 2018, 63(4): 51-52. [Hu Junjian. Guarding the desert and oasis: The record of ecological civilization construction of the 33rd regiment of the second division of Xinjiang Production and Construction Corps[J]. China Agricultural Reclamation, 2018, 63(4): 51-52.]

[11] 古丽努尔·沙布尔哈孜, 尹林克, 热合木都拉·阿地拉, 等. 塔里木河中下游退耕还林还草综合生态效益评价——以新疆生产建设兵团农二师33团为例[J]. 干旱区研究, 2004, 21(2): 161-165. [Gulnur Sabirhazi, Yin Linke, Rahmuttula Adil, et al. Evaluation on the compositive ecological benefits of withdrawing from farming to afforesting and grass planting in the middle and lower reaches of the Tarim River: A case study in 33rd regiment, second division, Xinjiang Group Company of Production and Construction [J]. Arid Zone Research, 2004, 21(2): 161-165.]

[12] 鲍士旦. 土壤农化分析[M]. 北京. 中国农业出版社, 2016. [Bao Shidan. Soil Agrochemical Analysis[M]. Beijing: China Agricultural Press, 2016.]

[13 杨晓武. 定边县板凳滩土壤系统分类与土壤肥力评价研究[D]. 杨凌: 西北农林科技大学, 2012. [Yang Xiaowu. Soil Taxonomy and Fertility Evaluation of Bandeng Bottomland in Dingbian Country[D]. Yangling: Northwest A & F University, 2012.]

[14] 张珊珊, 毛云玲, 冯志伟, 等. 云南松林下干巴菌生长土壤有效态微量元素特征及其影响因素[J]. 东北林业大学学报, 2021, 49(3): 108-112, 125. [Zhang Shanshan, Mao Yunling, Feng Zhiwei, et al. Available microelement characteristics of soil and its influential factors for the growth of *Thelephora ganbajun* under *Pinus yunnanensis* forest[J]. Journal of Northeast Forestry University, 2021, 49(3): 108-112, 125.]

[15] 桑慧茹, 王丽学, 陈韶明, 等. 基于主成分分析的RBF神经网络在需水预测中的应用[J]. 水电能源科学, 2017, 35(7): 58-61. [Sang Huiru, Wang Lixue, Chen Shaoming, et al. Water demand forecast model of RBF neutral networks based on principle component analysis[J]. Hydropower Energy Science, 2017, 35(7): 58-61.]

[16] 李良厚, 李吉跃. 聚类分析在立地分类与土壤肥力评价中的应用[C]// 第三届教育技术与培训国际会议论文集(第5卷), 湖北工业大学, 智能信息技术应用学会, 2010: 484-487. [Li Lianghou, Li Jiyue. Application of Clustering Analysis in Classifying Site Type and Evaluating Soil Fertility[C]// Proceedings of 2010 Third International Conference on Education Technology and Training(Volume 5), Hubei University of Technology, Intelligent Information Technology Application Association, 2010: 484-487.]

[17] 杨俊闯, 赵超. K-Means聚类算法研究综述[J]. 计算机工程与应用, 2019, 55(23): 7-14, 63. [Yang Junchuang, Zhao Chao. Survey on K-means clustering algorithm[J]. Computer Engineering and Applications, 2019, 55(23): 7-14, 63.]

[18] 刘雪娇. 数据挖掘中的动态聚类及增量研究[D]. 哈尔滨: 哈尔滨理工大学, 2015. [Liu Xuejiao. Research on Dynamic Clustering and Incremental in Data Mining[D]. Harbin: Harbin University of Science and Technology, 2015.]

[19] 陈迎丽, 何钰, 龚会琴, 等. 基于欧式距离法或因子化法的近红外光谱技术对牛肉掺假鉴定的研究[J]. 食品研究与开发, 2019, 40(15): 141-146. [Chen Yingli, He Yu, Gong Huiqin, et al. Study on identification of beef adulteration by near infrared spectroscopy based on euclidean distance method or factorization method[J]. Food Research and Development, 2019, 40(15): 141-146.]

[20] Saro, Kavita. Review: Study on simple k-mean and modified k-mean clustering technique[J]. International Journal of Computer Science Engineering and Technology, 2016, 6(7): 279-281.

[21] Anil K J, Data clustering: 50 years beyond k-means[J]. Pattern Recognition Letters, 2010, 31(8): 651-666.

[22] Hung C H, Chiou H M, Yang W N. Candidate groups search for K-harmonic means data clustering[J]. Applied Mathematical Modelling, 2013, 37(24): 10123-10128.

[23] 唐东凯, 王红梅, 胡明, 等. 优化初始聚类中心的改进K-means算法[J]. 小型微型计算机系统, 2018, 39(8): 1819-1823. [Tang Dongkai, Wang Hongmei, Hu Ming, et al. Optimizing initial cluster center of improved K-means algorithm[J]. Journal of Chinese Computer Systems, 2018, 39(8): 1819-1823.]

[24] 杨俊闯, 赵超. K-Means聚类算法研究综述[J]. 计算机工程与应用, 2019, 55(23): 7-14, 63. [Yang Junchao, Zhao Chao. Survey on K-means clustering algorithm[J]. Computer Engineering and Applications, 2019, 55(23): 7-14, 63.]

[25] 付亚丽, 李宏光, 付国润, 等. 红河植烟土壤中微量元素含量分析[J]. 云南农业大学学报(自然科学版), 2012, 27(1): 73-79. [Fu Yali, Li Hongguang, Fu Guorun, et al. Analysis of soil trace el-

ements in Honghe tobacco-growing area[J]. Journal of Yunnan Agricultural University(Natural Science), 2012, 27(1): 73−79.]

[26] 胡玲, 周丽娟, 王娟, 等. 云南烟区植烟土壤养分状况综合评价[J]. 河南农业科学, 2014, 43(7): 52−59. [Hu Ling, Zhou Lijuan, Wang Juan, et al. Comprehensive evaluation of soil fertility in tobacco-growing areas in Yunnan Province[J]. Journal of Henan Agricultural Sciences, 2014, 43(7): 52−59.]

[27] 陈江华. 中国植烟土壤及烟草养分综合管理[M]. 北京: 科学出版社, 2008. [Chen Jianghua. Integrated Management of Tobacco Soil and Tobacco Nutrients in China[M]. Beijing: Science Press, 2008.]

[28] 徐建华. 计量地理学[M]. 第一版. 北京: 高等教育出版, 2006: 81− 83. [Xu Jianhua. Quantitative Geography[M]. First Edition. Beijing: Higher Education Press, 2006: 81−83.]

[29] 陈留美, 桂林国, 吕家珑, 等. 应用主成分分析和聚类分析评价不同施肥处理条件下新垦淡灰钙土土壤肥力质量[J]. 土壤, 2008, 40(6): 971− 975. [Chen Liumei, Gui Linguo, Lyu Jialong, et al. Evaluation on soil fertility quality of newly cultivated light sierozem under different fertilization with methods of principal component and cluster analyses[J]. Soils, 2008, 40(6): 971−975.]

[30] 白由路, 金继运, 杨俐苹, 等. 农田土壤养分变异与施肥推荐[J]. 植物营养与肥料学报, 2001, 7(2): 129−133. [Bai Youlu, Jin Jiyun, Yang Liping, et al. Variability of soil nutrients in field and fertilizer recommendation[J]. Journal of Plant Nutrition and Fertilizers, 2001, 7(2): 129−133.]

[31] 刘继明, 宋启亮, 李芝茹, 等. 大兴安岭白桦低质林生态功能评价指标的灰色关联聚类分析[J]. 东北林业大学学报, 2012, 40(8): 112−115. [Liu Jiming, Song Qiliang, LI Zhiru, et al. Grey relation clustering analysis of evaluation indices of ecological functions of low-quality birch forests in greater Xing'an mountains after reconstruction by different transformation methods[J]. Journal of Northeast Forestry University, 2012, 40(8): 112−115.]

[32] 张新要, 李天福, 蒲文宣. 不同有机质含量土壤饼肥用量对烤烟产量及品质的影响[J]. 耕作与栽培, 2015, 35(5): 24−26. [Zhang Xinyao, Li Tianfu, Pu Wenxuan. Effects of different cake fertilizer amount on yield and quality of flue-cured tobacco in soil of different organic matter contents[J]. Tillage and Cultivation, 2015, 35(5): 24−26.]

[33] 许自成, 王林, 王金平. 湖南烤烟化学成分与土壤有机质含量的关系[J]. 生态学杂志, 2006, 25(10): 1186−1190. [Xu Zicheng, Wang Lin, Wang Jinping. Relationships between chemical components of flue-cured tobacco leaf and soil organic matter content in Hunan Province of China Chinese Journal of Ecology[J] Chinese Journal of Ecology, 2006, 25(10): 1186−1190.]

[34] 黄安, 杨联安, 杜挺, 等. 基于主成分分析的土壤养分综合评价[J]. 干旱区研究, 2014, 31(5): 819−825. [Hang An, Yang Lian'an, Du Ting, et al. Comprehensive assessment of soil nutrients based on PCA[J]. Arid Zone Research, 2014, 31(5): 819−825.]

[35] 田立文, 祁永春, 戴路, 等. 新疆南疆耕地土壤养分含量及其分布特征评价——以阿克苏地区为例[J]. 核农学报, 2019, 34(1): 214−223. [Tian Liwen, Qi Yongchun, Dai Lu, et al. Evaluation of soil nutrient content and its distribution of cultivated land in south of Xinjiang: Taking Aksu Prefecture as an example[J]. Journal of Nuclear Agricultural Sciences, 2019, 34(1): 214−223.]

[36] 袁宇尧, 李颖德, 任宇强, 等. 新疆阿拉尔垦区密植枣园土壤养分含量调查与分析[J]. 塔里木大学学报, 2016, 28(3): 1−6. [Yuan Yuyao, Li Yingde, Ren Yuqiang, et al. Survey and analysis of the soil nutrient in close planting jujube gardens in Alar area[J]. Journal of Tarim University, 2016, 28(3): 1−6.]

[37] 张丹, 罗格平, 许文强, 等. 新疆耕地土壤养分时空变化[J]. 干旱区地理, 2008, 31(2): 254−263. [Zhang Dan, Luo Geping, Xu Wenqiang, et al. Spatial-temporal change of soil nutrients of cultivated land in Xinjiang[J]. Arid Land Geography, 2008, 31(2): 254−263.]

[38] 陈彦. 绿洲农田土壤养分时空变异及精确分区管理研究[D]. 石河子: 石河子大学, 2008. [Chen Yan. Study on Spatio-temporal Variability and Definition of Management Zones of Soil Nutrients in Oasis field[D]. Shihezi: Shihezi University, 2008.]

[39] 孙亚洲. 基于分类距离的土壤系统分类土族单元划分研究——以河南省境内分布的主要土壤为例[D]. 郑州: 郑州大学, 2017. [Sun Yazhou. Classification of Soil Family based on Chinese Soil Taxonomy by Using Soil Taxonomic Distance: A Case Study of Soil in Henan province[D]. Zhengzhou: Zhengzhou University, 2017.]

Soil nutrient evaluation of cotton field based on distance clustering and K-means dynamic clustering

FAN Xianglong[1,2], LYU Xin[1,2], ZHANG Ze[1,2], GAO Pan[2,3],
ZHANG Qiang[1,2], YIN Caixia[1,2], YI Xiang[1,2]

(1. Key Laboratory of Oasis Eco-Agriculture, Xinjiang Production and Construction Group, Agriculture College of Shihezi Universit, Shihezi 832003, Xinjiang, China; 2. National and Local Joint Engineering Research Center for Agricultural Big Data of Xinjiang Production and Construction Corps, Shihezi 832003, Xinjiang; 3. School of Information Science and Technology, Shihezi University, Shihezi 832003, China)

Abstract: This study used principal component analysis combined with distance clustering, K-means dynamic clustering, and comprehensive soil nutrient evaluation methods to analyze and evaluate the soil nutrient status of cotton field soil in Xinjiang, China. The results showed that the available copper, manganese, and alkali hydrolyzable nitrogen played a major role principal component analysis: The available copper content was high (1.82 $mg \cdot kg^{-1}$); the available manganese content was low (11.36 $mg \cdot kg^{-1}$); and the alkali hydrolyzable nitrogen content was 122.07 $mg \cdot kg^{-1}$. The available phosphorus, potassium, zinc and iron contents were low. Thus, the distribution of soil nutrients was uneven. In the distance clustering and K-means dynamic clustering, the organic matter, available nitrogen, and available manganese contents were lower, whereas the other nutrient contents were higher. In the distance cluster, the soil nutrients could be expressed as: class I > class V > class IV > class II > class III, whereas the soil nutrients could be expressed in the K-means dynamic cluster as: class III > class I > class V > class II > class IV. In the comprehensive evaluation of soil nutrients (IFI), the grades of 1 and 16 are higher. The second company, third company, fourth company, sixth company, 15th company, 19th company, and the second prison area had higher levels, whereas the 8th, 9th, 10th, 11th, 12th, 17th, 18th and 20th companies were in the middle level. The grades of companies 5, 7, and 1 supervision district and agricultural city station were lower. Thus, K-means dynamic clustering is better than distance clustering, as it can be more scientific, reasonable, accurate, and effective for the comprehensive evaluation of soil nutrients.

Keywords: soil nutrients; distance clustering; K-means dynamic clustering; comprehensive evaluation value